

# Convex Hulls of Multiple Random Walks: A Large-Deviation Study

Timo Dewenter,<sup>1</sup> Gunnar Claussen,<sup>1,2</sup> Alexander K. Hartmann,<sup>1,\*</sup> and Satya N. Majumdar<sup>3,†</sup>

<sup>1</sup>*Institut für Physik, Universität Oldenburg, D-26111 Oldenburg, Germany*

<sup>2</sup>*Fachbereich Ingenieurwissenschaften, Jade Hochschule Wilhelmshaven/Oldenburg/Elsfleth, D-26389 Wilhelmshaven, Germany*

<sup>3</sup>*Laboratoire de Physique Théorique et Modèles Statistiques (UMR 8626 du CNRS),  
Université de Paris-Sud, Bâtiment 100, 91405 Orsay Cedex, France*

(Dated: May 24, 2016)

We study the polygons governing the convex hull of a point set created by the steps of  $n$  independent two-dimensional random walkers. Each such walk consists of  $T$  discrete time steps, where  $x$  and  $y$  increments are i.i.d. Gaussian. We analyze area  $A$  and perimeter  $L$  of the convex hulls. We obtain probability densities for these two quantities over a large range of the support by using a large-deviation approach allowing us to study densities below  $10^{-900}$ . We find that the densities exhibit a universal scaling behavior as a function of  $A/T$  and  $L/\sqrt{T}$ , respectively. As in the case of one walker ( $n = 1$ ), the densities follow Gaussian distributions for  $L$  and  $\sqrt{A}$ , respectively. We also obtained the rate functions for the area and perimeter, rescaled with the scaling behavior of their maximum possible values, and found limiting functions for  $T \rightarrow \infty$ , revealing that the densities follow the large-deviation principle. These rate functions can be described by a power law for  $n \rightarrow \infty$  as found in the  $n = 1$  case. We also investigated the behavior of the averages as a function of the number of walks  $n$  and found good agreement with the predicted behavior.

PACS numbers: 02.50.-r, 75.40.Mg, 89.75.Da

## I. INTRODUCTION

Originally, random walks have been introduced by Pólya [1] in 1921. Since then, many studies have dealt with this topic, as they are an ubiquitous model for physical, biological and social processes [2–4]. Example applications from biology include self-propelled motion of bacteria, and the diffusion of nutrients [3], as well as animal motion in general [5, 6]. Another example is the marking of territories by animals or the description of home ranges [7–9]. For the latter case a strong increase after the introduction of automated radio/GPS tagging of animals [10, 11]. The usage of minimum convex polygons, called *convex hulls*, bordering the trace of an animal [7, 12] is a simple yet versatile [13] way to describe the home range and can be used for any type of (random-walk) data. In two dimensions, the convex hull of a point set is the minimum subset whose elements form a convex polygon in such a way that (a) all points of the set and (b) the connecting lines between all possible pairs lie inside the polygon.

Much progress has been made on the analytical side, when the number of steps is very large and the random walk (with a finite variance of the step size) converges to the continuous-time Brownian motion (for a review see e.g., [14]). The mean perimeter [15, 16] and mean area [17] of a single two-dimensional Brownian motion are known for a long time.

It was shown [14, 18] recently that the problem of com-

puting the mean perimeter and the mean area of the convex hull of an arbitrary two-dimensional stochastic process can be mapped to computing the extremal statistics of the one-dimensional component of the process. This procedure was successfully applied recently to compute the mean perimeter and the mean area of several two-dimensional stochastic processes such as the random acceleration process in 2D [19], 2D branching Brownian motions with absorption and applications to epidemic outbreak [20] and 2D anomalous diffusion processes [21]. Very recently, this method was also successfully used to compute the exact mean perimeter of the convex hull of a planar Brownian motion confined to a half-space [22]. Finally, using different methods, the mean perimeter and the mean area of the convex hull of a single Brownian motion, but in arbitrary dimensions, have been computed recently in the mathematics literature [23, 24].

Analytical calculations of even the second moment for the area and perimeter of a convex hull regarding single two-dimensional Brownian motion turned out to be very difficult [25, 26]. For the full distributions of the area and perimeter no analytical results are known so far, so the usage of computer simulations is a natural approach, as done in a recent study [27].

Here, we are interested in multiple random walkers, which perform their walks independently from each other. The investigation of  $n$  non-interacting random walkers on a  $d$ -dimensional regular lattice has been done in [28–30]. Many studies have been published for interacting multiparticle walkers [31], e.g., in one dimension [32–39]. The mean first passage time of  $n$  independent diffusing particles in Euclidean space is calculated, e.g., in [40] or [41]. Recently, the mean perimeter and the mean area of  $n$  independent Brownian motions have been computed in 2D [14, 18]. In this article we want to check

\*a.hartmann@uni-oldenburg.de; <http://www.compphys.uni-oldenburg.de/en>

†satya.majumdar@u-psud.fr

the predictions from theory by numerical simulations of  $n$  non-interacting time-discrete Brownian random walkers as well as the probability density functions of the area and perimeter of the corresponding convex hulls. In particular we apply a numerical large-deviation approach to obtain the probability density functions over a large range of the support, down to probability densities as small as  $10^{-900}$ . In addition, we are interested in what way the same results for  $n > 1$  walkers are found in comparison to the  $n = 1$  case [27].

The paper is organized as follows: Section II introduces the random walk model, the convex hull of a two-dimensional point set, and briefly elucidates an algorithm to obtain such a convex hull. Part III explains the large-deviation scheme used to obtain the probability density function over a large range of the support including the low-probability tails. The next part IV presents the results achieved from our simulations. The last section V concludes the article and a short outlook is given.

## II. RANDOM WALKS, CONVEX HULLS, AND ALGORITHMS

A time-discretized random walk consists of  $T$  step vectors  $\vec{\delta}_i$ , and the position  $\vec{x}(\tau)$  at time step  $\tau < T$  is the sum of all steps up to  $\tau$ , i.e.:

$$\vec{x}(\tau) = \vec{x}_0 + \sum_{i=1}^{\tau} \vec{\delta}_i. \quad (1)$$

The walk configuration itself is then the set  $\mathcal{W} = \{\vec{\delta}_1, \vec{\delta}_2, \dots, \vec{\delta}_T\}$  of steps [42]. The step  $\vec{\delta}_i = (\delta_{x,i}, \delta_{y,i})$  itself denotes a displacement of the particle by  $\delta_{x,i}$  in  $x$ -direction and  $\delta_{y,i}$  in  $y$ -direction. Here, we consider a time-discrete approximation to a Brownian walk, i.e., both  $\delta_{x,i}$  and  $\delta_{y,i}$  are, for each  $i$ , drawn randomly from a Gaussian distribution with zero mean and variance one. All considered walks are open, i.e., the walker does not need to get back to the starting point  $\vec{x}(0)$  after  $T$  steps.

In contrast to [27], where only single walks with one walker have been investigated, we put multiple random walks under scrutiny. So, starting from the origin of the coordinate system,  $n$  independent random walkers perform their walks simultaneously. The resulting point set  $\mathcal{W}$  of  $n \cdot T$  points given by the individual positions of all  $n$  walkers after each time step is then further investigated.

The convex hull  $\mathcal{C} = \text{conv}(\tilde{\mathcal{P}})$  of a two-dimensional point set  $\tilde{\mathcal{P}} = \{\tilde{P}_i\}$ ,  $\tilde{P}_i \in \mathbb{R}^2$  is described through a convex set over  $\tilde{\mathcal{P}}$ . The points  $P$  within  $\mathcal{C}$  are given by all possible combinations  $P = \sum_i \alpha_i \tilde{P}_i$  with  $\tilde{P}_i \in \tilde{\mathcal{P}}$  and  $\sum_i \alpha_i = 1$  and  $\alpha_i \in \mathbb{R}_0^+$  (definition given according to [43]). This means:

1. All points  $\tilde{P}_i \in \tilde{\mathcal{P}}$  lie within  $\mathcal{C}$ .
2. All lines  $\overline{\tilde{P}_i \tilde{P}_j}$ ;  $\tilde{P}_i, \tilde{P}_j \in \tilde{\mathcal{P}}$  also lie within  $\mathcal{C}$ .

The boundary of the convex set is a polygon which connects a subset  $\mathcal{P} \subset \tilde{\mathcal{P}}$  of  $H$  points from the point set, i.e.,  $\mathcal{P} = \{P_0, P_1, \dots, P_{H-1}\}$ , with  $P_i = (x_i, y_i)$  ( $i = 0, \dots, H-1$ ). The hull is attributed with area  $A$  and perimeter  $L$  according to (identifying  $i = H$  with  $i = 0$ ):

$$A(\mathcal{C}) = \frac{1}{2} \sum_{i=0}^{H-1} (y_i + y_{i+1})(x_i - x_{i+1}) \quad (2)$$

$$L(\mathcal{C}) = \sum_{i=0}^{H-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \quad (3)$$

For our work, we determined the polygons bordering convex hulls (for which one uses shortly the term “convex hull”) numerically. For convenience, we use dimensionless quantities subsequently, as all convex hulls are represented in a computer.

Here, we used the “Jarvis March” algorithm [44], which has a complexity of  $\mathcal{O}(N \cdot H)$ , where  $N$  is the number of points in the investigated point set and  $H$  the number of points in the convex hull. In this algorithm, the convex hull is calculated in a “gift-wrapping” manner, where one needs to make sure that all points of the set lie on e.g., the right side of a starting point. The next point added to the convex hull is the point which has the minimum angle between the line connecting both points and the vertical. This procedure is repeated until one reaches the starting point again.

In usual cases, the application of convex hull algorithms can be accelerated by usage of pre-selection heuristics, such as the one introduced by Akl and Toussaint [45]. This heuristic looks up extreme points of the set (i.e., those of maximum and minimum  $x$ - and  $y$ -coordinates) and discards all points which lie inside the quadrilateral formed by these points. We use a custom refinement of this heuristic, which is based on iterating the heuristic under rotation of the coordinate origin, which eliminates another fraction of inert points per each iteration.

## III. LARGE-DEVIATION SCHEME

For simple-sampling results, walk configurations  $\mathcal{W}$  for multiple walkers  $n$  are generated randomly, and the according convex hulls  $\mathcal{C}$  are calculated through the algorithm, resulting in a multitude of values of  $A$  and  $L$ . Obtaining histograms of these values only gives access to the high probability regime, where the convex-hull properties of typical random walks are measured. However, in order to obtain values of these quantities with especially low probabilities, allowing us to measure the distributions  $P_n(A)$  and  $P_n(L)$  over a large range of the support, a certain Markov-Chain Monte Carlo (MCMC) scheme can be used [46, 47].

The MCMC consists of an evolution of random walks  $\mathcal{W}(t)$  and corresponding sets  $\tilde{\mathcal{W}}(t)$  of points.  $t$  is another discrete “time” parameter, not to be confused with the

time parameter  $\tau$  of the random walks. For the walks, we measure the property  $S(t)$ , i.e., the area ( $S = A$ ) or perimeter ( $S = L$ ) of the convex hull of the point sets, depending on which distribution  $P_n(A)$  or  $P_n(L)$  we are aiming at. The initial configuration  $\mathcal{W}(0)$  is any walk configuration, e.g., a randomly chosen one.

At each Monte Carlo step  $t$ , all  $n$  independent walks  $\mathcal{W}_k(t)$  ( $k \in \{1, 2, \dots, n\}$ ) are altered to  $\mathcal{W}_k^*$  by replacing one randomly selected step  $\vec{\delta}_i$  ( $i \in \{1, 2, \dots, T\}$ ) in each walk with a newly generated step  $\vec{\delta}_i'$ . The new step is generated according to the same distribution as all other random walk steps, i.e., the  $x$ - and  $y$ -coordinate of  $\vec{\delta}_i'$  are drawn independently from a Gaussian distribution. Note that by exchanging e.g., the first step  $\vec{\delta}_1$ , all following positions  $\vec{x}$  (cf., Eq. (1)) of the walk are changed. The convex hull of the point set  $\tilde{\mathcal{W}}^* = \bigcup_k \tilde{\mathcal{W}}_k^*$  resulting from the  $n$  walks  $\mathcal{W}_k^*$  is calculated, leading to the quantity  $S^*$ . The alteration  $\mathcal{W}^*$  is *accepted* ( $\mathcal{W}(t+1) = \mathcal{W}^*$ ) according to the Metropolis probability:

$$p_{\text{Met}} = \min \left[ 1, e^{-(S^* - S(t))/\Theta} \right]. \quad (4)$$

Here,  $\Theta$  is the (artificial) Monte Carlo “temperature”, which is a parameter used to set the range of the sampled values. If the alteration is not accepted, it is *rejected*, i.e.,  $\mathcal{W}(t+1) = \mathcal{W}(t)$ .

Like in any MCMC simulation one needs to equilibrate the simulation, i.e., discards the initial part of the measured quantities until “typical” values are found. Typical equilibration times are  $10^3$  sweeps (one sweep equals  $T$  MC steps) for e.g.,  $T = 200$ ,  $n = 3$ , and  $\Theta = 10$  for the area of the convex hull. In addition, we pick only each  $k$ th data point from the original measurement to get roughly decorrelated values. For the case above we use  $k = 1$  sweeps, which is a typical value. Towards low absolute values of the temperature this value needs to be increased, so e.g., for  $\Theta = 0.2$  in the above case, we choose  $k = 100$  sweeps.

For a given quantity ( $S = A$  or  $S = L$ ) and a given walk length  $T$  one gets different probability density functions (pdfs)  $P_\Theta(S)$  for each temperature  $\Theta$  used. They are related to the actual distribution  $P(S)$  according to the relation [46]

$$P(S) = e^{S/\Theta} Z(\Theta) P_\Theta(S), \quad (5)$$

where  $Z(\Theta)$  is a normalization constant. For different values of  $\Theta$ , different ranges of the measured value  $S$  are obtained. This allows for a piecewise reconstruction of  $P(S)$  via suitable choices of the normalization constants  $Z(\Theta)$ . They can be calculated through inversion of this formula whenever for two values  $\Theta_1$  and  $\Theta_2$  the ranges of the sampled values of  $S$  overlap. Thus, the temperatures are chosen such that for neighbouring  $\Theta$  the measured histograms sufficiently overlap.

For a more detailed description of the calculation of the normalization constants  $Z(\Theta)$  and the determination

of the pdf from the pdfs for the single temperatures we refer to, e.g., [46, 48].

Note that the large-deviation approach has already been applied successfully for the case of the convex hull of the point set of one ( $n = 1$ ) walker [27]. In that reference also the test case of independent points was simulated and a comparison with analytical results yielded a good agreement.

## IV. RESULTS

For  $n = 3$  random walks we performed simulations for walk lengths of  $T \in [20, 200]$  while measuring and biasing for the area  $A$  and the perimeter  $L$  of the convex hulls, respectively. To obtain a large range of the support for the pdfs of these two quantities we used e.g. 17 temperatures  $\Theta \in [-200, 40]$  (excluding the value  $\Theta = \infty$ , which corresponds to simple sampling) for  $T = 200$  for the pdf of  $A$  and about 40 temperatures  $\Theta \in [-20, 5]$  for  $L$ .

We also studied the case of  $n = 2$  walks, which is closer to the single walker case. Here, we used walks of lengths  $T \in [20, 500]$ .

To investigate the behavior with increasing number of walks  $n$ , due to the strongly increasing numerical effort, we performed simulations at fixed system size  $T = 50$  and variable number of walks  $n \in [2, 6]$  for both observables  $A$  and  $L$ . We again obtained probability density functions over a large range of the support. In addition, we performed simple-sample simulations, i.e., close to the peak of the histogram, for  $T = 50$ ,  $10^6$  samples and up to  $n = 10^5$  independent random walkers.

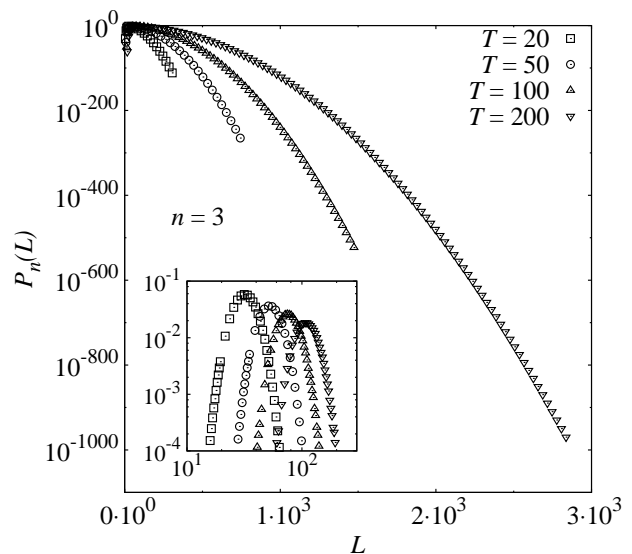


FIG. 1: Probability density function  $P_n(L)$  of the perimeter  $L$  of the convex hull of  $n = 3$  independent random walks in semi-logarithmic scale. Inset: Region around peaks in double logarithmic scale.

### A. Probability density function

As an example, Fig. 1 shows the pdf of the perimeter of the convex hull of three independent two-dimensional time-discrete open Brownian walks. By using the large-deviation approach, probability densities smaller than  $10^{-900}$  can be reached. One can observe the strong curvature of the data on a semi-logarithmic scale. With increasing walk length  $T$  the probability densities also increase when looking at a fixed perimeter. This is due to the fact that for larger walk lengths large perimeters are found by the simulations more likely as more steps in the random walk are available. We obtained results with similar high numerical quality for the probability density of the area (not shown without rescaling for brevity).

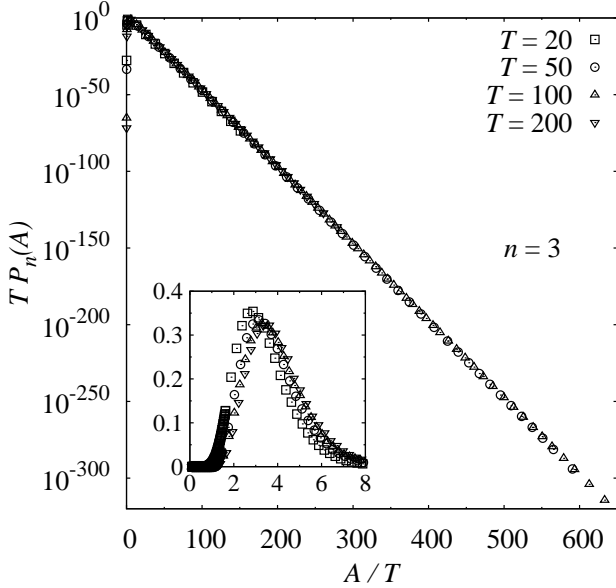


FIG. 2: Rescaled pdfs for  $n = 3$  walks, the area  $A$  of the convex hull and different walk lengths  $T$  in semi-logarithmic scale. Inset: Region close to peaks in linear scale.

Next, we check whether the scaling assumptions for the area [14]

$$P_{T,n}(A) = \frac{1}{T} \tilde{P}_n\left(\frac{A}{T}\right), \quad (6)$$

and the perimeter [14]

$$P_{T,n}(L) = \frac{1}{\sqrt{T}} \tilde{P}_n\left(\frac{L}{\sqrt{T}}\right), \quad (7)$$

are also valid [27] in the case of multiple ( $n = 2, 3$ ) random walks. Here  $\tilde{P}_n(\cdot)$  are universal distributions (actually different ones, here distinguished by the argument  $A$  and  $L$ , respectively) independent of  $T$ . This scaling behavior represents the known scaling of the mean values as function of walk length  $T$ .

In Fig. 2 the collapse according to Eq. (6) is shown. In the tail of the rescaled pdfs almost perfect agreement of

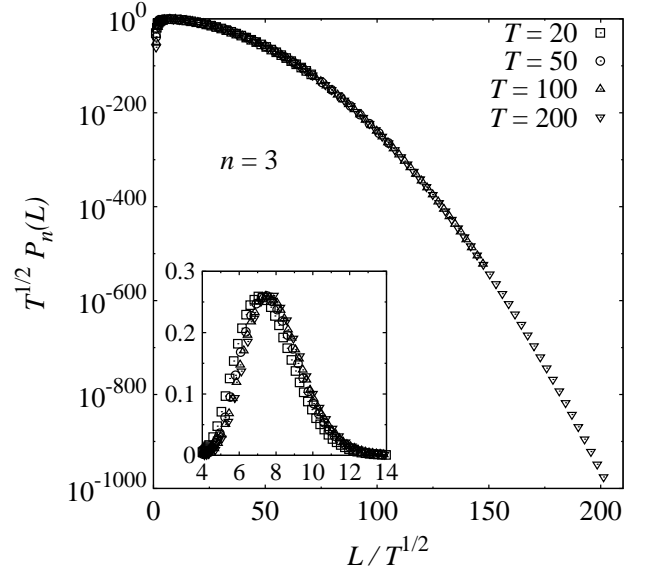


FIG. 3: Rescaled pdfs for  $n = 3$  walks, the perimeter  $L$  of the convex hull and different walk lengths  $T$  in semi-logarithmic scale. Inset: Region close to peaks in linear scale.

the curves for the different system sizes is visible. Only in the peak region (cf. inset of Fig. 2) small finite-size effects occur. The collapse of the pdfs for the perimeter in accordance with Eq. (7) is depicted in Fig. 3. A good collapse with small finite-size deviations in the peak region (see inset) is also achieved.

Similar results were found and a good data collapse was achieved (not shown) for the area and perimeter for  $n = 2$ , respectively.

### B. Functional form of the probability density function

According to [27] we use as universal distributions  $\tilde{P}_n(\cdot)$  in Eqs. (6) and (7) two Gaussians with mean  $\mu_S$  and standard deviation  $\sigma_S$  ( $S = L$  or  $S = A$ ) in the case of large  $T$ . For the perimeter we obtain [27]:

$$\tilde{P}_n(m) = \frac{a_L}{\sqrt{2\pi\sigma_L^2}} \exp\left(-\frac{(m - \mu_L)^2}{2\sigma_L^2}\right), \quad (8)$$

where  $a_L$  is a constant and  $m = L/\sqrt{T}$ . If we approximate  $A \propto L^2$ , so  $l \propto m^2$  with  $l = A/T$ , we get an additional factor  $1/\sqrt{l}$  from  $|dm/dl| \sim 1/\sqrt{l}$  in the scaling relation. In total, the scaling function for the area is given by [27]:

$$\tilde{P}_n(l) = \frac{a_A}{\sqrt{2\pi\sigma_A^2 l}} \exp\left(-\frac{(\sqrt{l} - \mu_A)^2}{2\sigma_A^2}\right), \quad (9)$$

where  $a_A$  is some constant parameter.

Figs. 4 and 5 show the results of those fits to the rescaled pdfs for  $T = 200$  and  $n = 3$ . In both figures, two

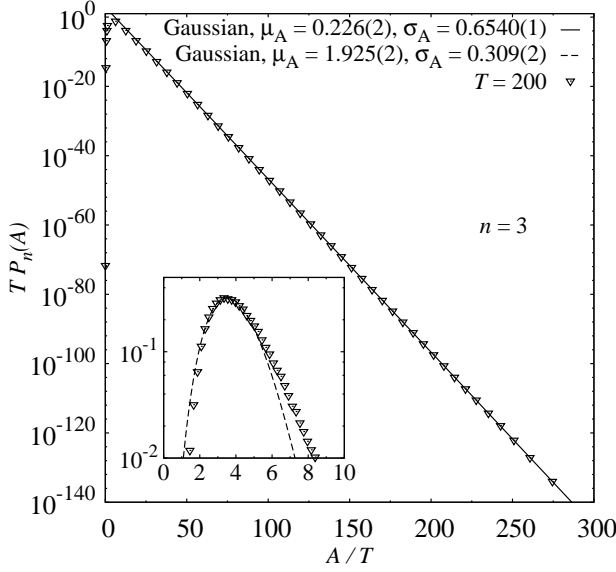


FIG. 4: Gaussian fit (solid line) according to Eq. (9) with parameter  $a_A = 1000(25)$  to the rescaled pdf of the area for  $T = 200$  and  $n = 3$ . Note that the logarithm of Eq. (9) was fitted to the logarithm of the pdf for  $A/T \geq 50$  to match the tail. Inset: Gaussian fit (dashed line) in the peak region for  $A/T \in [2, 5]$  corresponding to Eq. (9) with  $a_A = 0.474(3)$ .

independent fits, one to the tail of the distribution and one to the peak region, were necessary. A single fit over the full support does not match the data. In Fig. 4 both independent Gaussian fits fit the data well although deviations around the peak (cf., inset of Fig. 4) are visible.

Fig. 5 shows the fits according to Eq. (8) to the rescaled pdf of the perimeter. The fit to the tail matches the region of large  $L/\sqrt{T}$  very well. Nevertheless, in the inset of Fig. 5 strong deviations from a Gaussian behavior can be seen.

Next, we investigate the (left) tail of the pdfs towards small values of the rescaled area and perimeter. Corresponding to [27] we expect for the perimeter asymptotically for small  $m$  and small  $T$  an essential singularity according to

$$\tilde{P}_n(m) \sim a \exp\left(-\frac{b}{m^2}\right), \quad (10)$$

where  $a$  and  $b$  are constants and again  $m = L/\sqrt{T}$ . With similar arguments as for large  $T$ , where Gaussian fits are used (cf., Eqs. (8) and (9)) we obtain for the small  $l$  asymptotics of the area

$$\tilde{P}_n(l) \sim \frac{a}{\sqrt{l}} \exp\left(-\frac{b}{l}\right), \quad (11)$$

where  $a$  and  $b$  are constants and again  $l = A/T$ .

Fig. 6 shows the fits to the left tails of the rescaled pdfs of the area and perimeter, respectively. The fit to the rescaled pdf of the perimeter matches the data quite well for small  $m < 5$ . Also the fit according to Eq. (11)

(see inset of Fig. 6) suits well to the rescaled pdf of the area for small  $l < 2$ .

### C. Rate function

Next, the empirical rate function  $\Phi_n(s)$  [49] is calculated which describes the leading behavior of the pdf in the large-deviation tail. If one assumes that the behavior of the probability density away from the typical values around the peak is exponentially small in the walk length  $T$ , one gets for the rate function

$$\Phi_n(s) = -\frac{1}{T} \ln P_n(s). \quad (12)$$

The quantity  $s$  is usually normalized with the maximum possible values so that  $s \in [0, 1]$ . As for Gaussian random walks no real maximum exists, we choose [27]  $s_A = A/T^2$  and  $s_L = L/T$ , respectively.

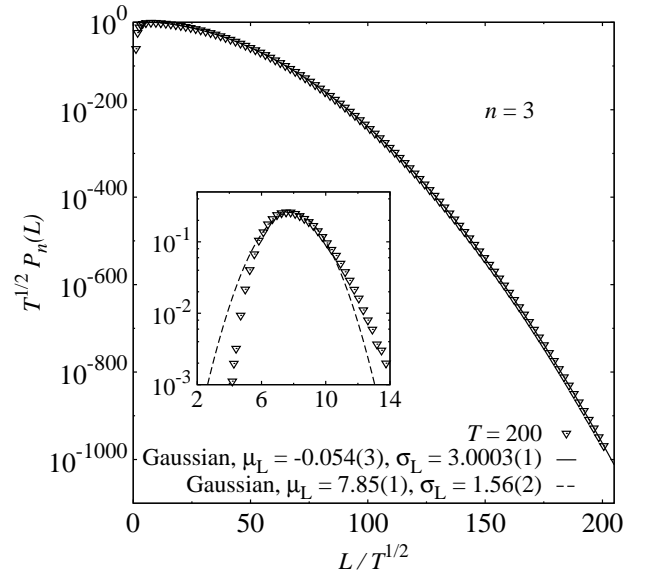


FIG. 5: Gaussian fit (solid line) according to Eq. (8) with parameter  $a_L = 1200(22)$  to the rescaled pdf of the perimeter for  $T = 200$  and  $n = 3$ . Note that the logarithm of Eq. (9) was fitted to the logarithm of the pdf for  $L/\sqrt{T} \geq 20$  to match the tail. Inset: Gaussian fit (dashed line) in the peak region for  $L/\sqrt{T} \in [5, 15]$  corresponding to Eq. (8) with  $a_L = 1.011(8)$ .

Fig. 7 shows the rate function for  $n = 3$  and the area of the convex hull. For small values of  $s_A$  there are a strong finite length effects, whereas for larger values the curves for different  $T$  seem to converge quickly to one curve. Nevertheless, a convergence of the rate function to one universal shape seems likely, indicating that the densities obey the large-deviation principle [49].

To estimate the behavior of the curves for large  $T$  we plotted the rate function in a double-logarithmic scale in the inset of Fig. 7 and also show a power law  $s_A^\kappa$  with  $\kappa = 1$  for comparison. Apparently, our data has the same

slope, at least in the region where  $s_A$  is large. This is the same result as was found previously for  $n = 1$  [27].

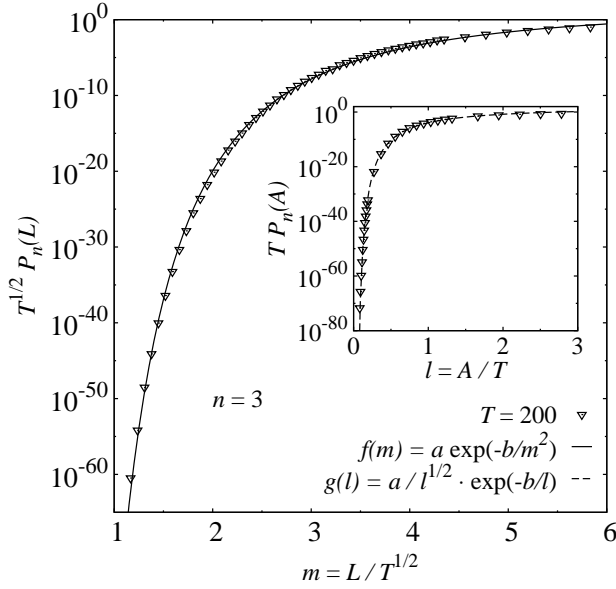


FIG. 6: Fit (solid line) with fit function  $f(m)$ , which corresponds to Eq. (10) to the rescaled pdf of the perimeter for  $T = 200$  and  $n = 3$ . Parameters of the fit are  $a = 72(8)$  and  $b = 200.9(5)$ . Note that the logarithm of the fit function was fitted to the logarithm of the pdf for  $m \in [1, 5]$  to match the tail of the pdf. Inset: Fit (dashed line) with fit function  $g(l)$  corresponding to Eq. (11) to the rescaled pdf of the area for  $T = 200$  and  $n = 3$ . Parameters of the fit are  $a = 394(22)$  and  $b = 15.06(2)$ . Note that the logarithm of the fit function was fitted to the logarithm of the pdf for  $l \in [0, 2]$  to match the tail of the pdf.

In Fig. 8 the rate function for the perimeter and  $n = 3$  is depicted. Again, for small values of  $s_L$  strong finite size effects occur and the convergence to a common curve is very slow. On the other hand, for larger  $s_L$  the convergence to a common curve is already visible. In the inset of Fig. 8 the data is shown in a double-logarithmic plot. Our data is compatible with a power-law behavior  $s_L^\kappa$  with  $\kappa = 2$  for large  $s_L$  represented by the dashed line. Again we have found the same result as was found previously for  $n = 1$  [27].

Thus, for few number of walks, the behavior of the rate functions for large  $s$  of both the area and the perimeter agree with the expected ones [27] found for  $n = 1$ . So, our pdfs are said to follow the “large-deviation principle” as they can be well described by a rate function given by Eq. (12). The behavior of the rate function when increasing the number of walks more strongly is discussed below in the following section.

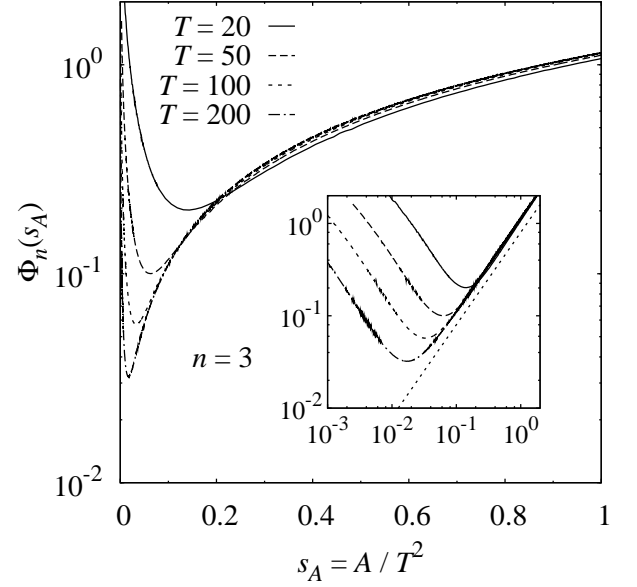


FIG. 7: Rate function  $\Phi_n(s_A)$  as a function of the scaled area  $s_A = A/T^2$  for different walk lengths  $T$  and  $n = 3$  walks in semi-logarithmic scale. Inset: The same in a double-logarithmic plot, where the dashed line close to the data is a power law  $s_A^\kappa$ , with  $\kappa = 1$ .

#### D. Scaling behavior with respect to the number $n$ of walks

According to [14] for  $n$  two-dimensional open Brownian random walks the average area is expected to scale like

$$\langle A_n \rangle = \beta_n T, \quad (13)$$

where the  $n$ -dependent prefactor is given by

$$\beta_n = 4n\sqrt{\pi} \int_0^\infty u [\text{erf}(u)]^{n-1} \cdot [ue^{-u^2} - g(u)] du, \quad (14)$$

$\text{erf}(u)$  is the error function

$$\text{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt,$$

and

$$g(u) = \frac{1}{2\sqrt{\pi}} \int_0^1 \frac{e^{-u^2/t}}{\sqrt{t(1-t)}} dt.$$

In the large- $n$  limit  $\beta_n$  scales like [14]

$$\beta_n \sim 2\pi \ln n. \quad (15)$$

According to [14] the average perimeter of convex hulls of  $n$  two-dimensional Brownian walks should scale like

$$\langle L_n \rangle = \alpha_n \sqrt{T}, \quad (16)$$

with

$$\alpha_n = 4n\sqrt{2\pi} \int_0^\infty ue^{-u^2} [\text{erf}(u)]^{n-1} du, \quad (17)$$

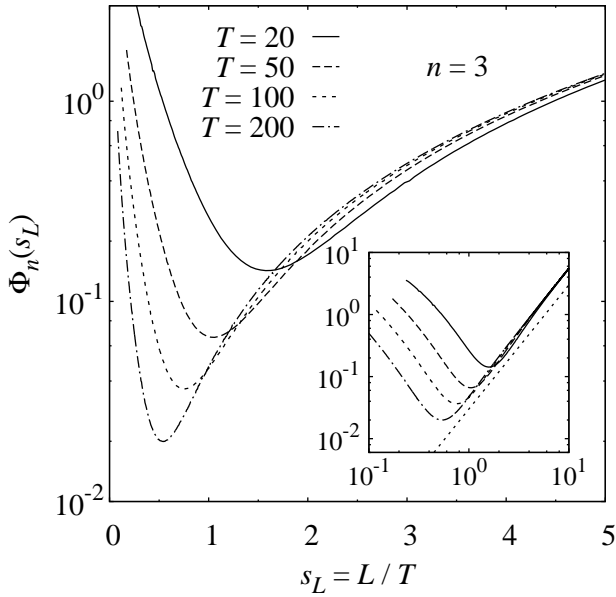


FIG. 8: Rate function  $\Phi_n(s_L)$  as a function of the scaled perimeter  $s_L = L/T$  for different walk lengths  $T$  and  $n = 3$  walks in semi-logarithmic scale. Inset: The same in a double-logarithmic plot, where the dashed line close to the data is a power law  $s_L^\kappa$ , with  $\kappa = 2$ .

which has a large- $n$  scaling

$$\alpha_n \sim 2\pi\sqrt{2 \ln n}. \quad (18)$$

To check these analytical predictions, we performed simple-sampling simulations to determine the average area and perimeter for various values of  $n \in [1, 10^5]$ . Fig. 9 shows the results for  $n = 3$  independent Gaussian walks and the rescaled averages  $\mu_A = \langle A \rangle / T$  and  $\mu_L = \langle L \rangle / \sqrt{T}$ , where  $\langle \cdot \rangle$  denotes averaging. We simulated walk lengths  $T \in [10, 2000]$  and used at least  $8 \cdot 10^5$  samples to determine the average. A power-law fit

$$\mu_S(L) = \mu_S^\infty + a \cdot T^b \quad (19)$$

with parameter  $\mu_S^\infty$  ( $S = A$  or  $S = L$ ) denoting the extrapolated value for  $T \rightarrow \infty$ , and fit parameters  $a, b$  is performed. Excluding the small system sizes the fit is done over the range  $T \in [150, 2000]$ , yielding a reduced chi-square value of  $\chi_{\text{red}}^2 \approx 0.72$  for the area. One can observe a convergence towards the average area for infinite  $T$ , which is  $\mu_A^\infty = 4.415(6)$ . Compared to the literature [14]  $\beta_3 = \pi + 3 - \sqrt{3} \approx 4.410$  the measured value agrees within error bars. The inset of Fig. 9 shows the fit to the rescaled average perimeter  $\mu_L$ . Again, a convergence towards the average for  $T \rightarrow \infty$  is visible. The fit gives  $\chi_{\text{red}}^2 \approx 0.76$  for  $T \in [150, 2000]$  with  $\mu_L^\infty = 8.339(7)$ . This value is compatible within error bars with the analytical derivation [14]  $\alpha_3 \approx 8.334$ .

For  $n = 2$  walks we accomplished similar fits (not shown) as for  $n = 3$ . The fit for the rescaled average area was performed with an reduced chi-square value

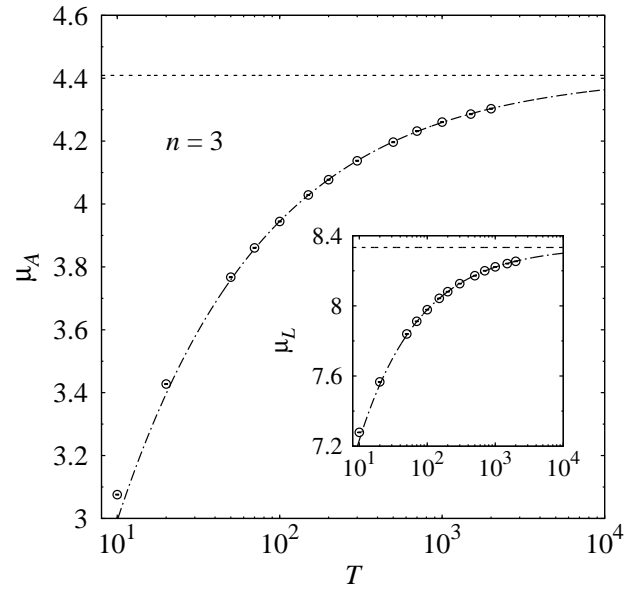


FIG. 9: Average rescaled area  $\mu_A$  of the convex hull as a function of walk length  $T$  for  $n = 3$  walks. Note the logarithmic scaling of the  $T$ -axis. For each data point at least  $8 \cdot 10^5$  samples were used. Dashed-dotted line is a power-law fit according to Eq. (19). Horizontal dashed line represents analytical expectation [14]  $\beta_3 \approx 4.410$ . Inset: The same for the average rescaled perimeter  $\mu_L$ . Horizontal line denotes expectation according to [14]:  $\alpha_3 \approx 8.334$ .

of  $\chi_{\text{red}}^2 \approx 0.20$  over system sizes  $T \in [300, 2000]$ . The average value for an infinite system is  $\mu_A^\infty = 3.144(5)$ , which is compatible within error bars with literature [14]:  $\beta_2 = \pi \approx 3.142$ . For the average perimeter we obtained by the fit according to Eq. (19)  $\chi_{\text{red}}^2 \approx 0.10$  with  $\mu_L^\infty = 7.091(4)$ . This value agrees within error bars with the published value [14]  $\alpha_2 = 4 \cdot \sqrt{\pi} \approx 7.090$ .

Next, we want to check if our data matches the exact equations (14) and (17) for high values of  $n$ . In Fig. 10 the averages of the area and perimeter obtained from simple-sampling simulations with  $T = 50$  fixed and various values for the number of walks  $n$  is presented. Scaling the  $n$ -axis logarithmically leads to a linear behavior of the average area for large  $n$  indicating a logarithmic dependence like expected by the previous scaling assumptions for  $\beta_n$ . As Eq. (14) is only valid for large values of  $T$  one can see a small deviation between the theoretical and the measured values.

In the same way, the behavior of the average perimeter follows the expected behavior, as shown in the inset of Fig. 10.

Clearly, the data points for  $T = 50$  are located systematically below the analytical curves, which is only valid for  $T \rightarrow \infty$ . This does not come unexpectedly, because we see this behavior already for  $n = 3$  in Fig. 9. To check the convergence of the data for different walk lengths we investigate (Figures not shown) convex hulls for  $n = 100$  walks as already done for  $n = 3$ . A fit according to Eq. (19) for walk lengths  $T \in [70, 10^4]$  yields for the area

a reduced chi-square value of  $\chi_{\text{red}}^2 \approx 1.4$ . One can observe a convergence for  $T \rightarrow \infty$  towards  $\mu_A^\infty = 21.40(1)$  which is compatible with the theoretical value (cf., Eq. (14))  $\beta_{100} \approx 21.3890$  within a standard error bar. For the perimeter the power-law fit for  $T \in [100, 10^4]$  results in  $\chi_{\text{red}}^2 \approx 0.18$  and an extrapolated value for infinite  $T$  which is  $\mu_L^\infty = 17.262(1)$ . Compared to the theoretical value  $\alpha_{100} \approx 17.2596$  there is good agreement within two standard error bars. So, we can be confident that our data shows the expected convergence for all values of  $n$  towards the theoretical values.

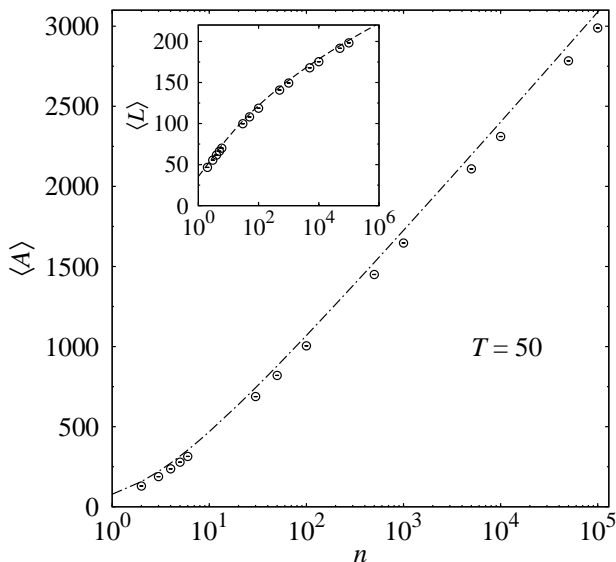


FIG. 10: Average area as a function of number of walks  $n$  for  $T = 50$ . Note the logarithmic scaling of the  $n$ -axis. For each data point about  $10^6$  samples have been used. The dashed-dotted line shows the exact value given by Eqs. (13) and (14). Inset: The same for the averaged perimeter. Dashed line displays exact value obtained by Eqs. (16) and (17).

Next, we want to investigate whether the scaling behavior of the average with respect to the number  $n$  of walks transfers to the full distributions, as it is the case with respect to the number  $T$  of steps in the walks. Figs. 11 and 12 show the distributions with a corresponding rescaling of the axis. Apparently the quality of the collapse is not very good but seems to get gradually better when making the number  $n$  of walks very large. This can be seen when looking at the insets of the figures, where the change of the distributions for  $n = 10^4 \rightarrow n = 10^5$  is rather small, compared to the change  $n = 2 \rightarrow n = 6$ . This corresponds to the just discussed behavior of the mean, where also strong scaling corrections at small number  $n$  of walks are visible. Thus, a convergence to the scaling form, at rather large values of  $n$ , appears likely.

Finally, we consider the shape of the distributions in the limit of a large number of walkers and long walk lengths. The distribution of the perimeter, on which we focus here, can be approximated [14] by the distribution

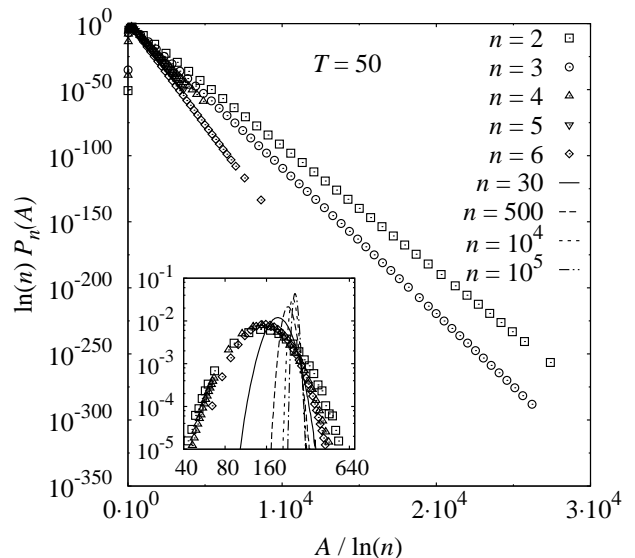


FIG. 11: Rescaled pdfs for  $T = 50$ , the area  $A$  of the convex hull and various number of walks  $n$  in semi-logarithmic scale. Inset: Region close to peaks in double-logarithmic scale. Note that for walk numbers  $n \geq 30$  only values from simple sampling exist and therefore only the region around the peak is depicted.

of  $2\pi$  times the *span* of  $n$  independent one-dimensional random walkers. Since the span is given by the sum of the two extreme points in positive and negative directions of the  $n$  one-dimensional walkers, this distribution is basically given by the convolution of two Gumbel extreme-value distributions. The exact distribution [50] for the one-dimensional case can be formulated in terms of the modified Bessel function  $K_0(x)$  of zero'th order. Correspondingly to Eq. (6) of Ref. [50] we fit our data for the perimeter to

$$f(L) = 2ab \exp(-z) K_0(2 \exp(-z/2)) \quad (20)$$

with  $b \equiv 2\sqrt{\log n}$ ,  $z \equiv b((L - B)/C - b)$ . The variables  $B$ ,  $C$  and  $a$  are fit parameters allowing for an adjustment of the center and width of the distributions, and taking care of the normalization, respectively.

In Fig. 13 the data for  $n = 1000$  and  $T = 10^4$  as obtained from simple sampling is shown together with a corresponding fit. The fit is good in the center of the distribution, but not away from it. Nevertheless, as the inset shows, when increasing either  $n$  or  $T$  the quality of the fit increases considerably.

Furthermore, we investigated the behavior of the rate function with varying number of walks  $n$ . We show  $\Phi_n(s_L)$  in Fig. 14 as a function of  $s_L = L/T$ , where  $T = 50$  is fixed, for various values of  $n$ , while using here again also the large-deviation data. For small  $s_L$  a strong influence of  $n$  can be seen, while this is weaker for larger values of  $s_L$ .

In the inset of Fig. 14 we show the data of Eq. (20) with the parameters as obtained from the fit for  $n = 1000$  and



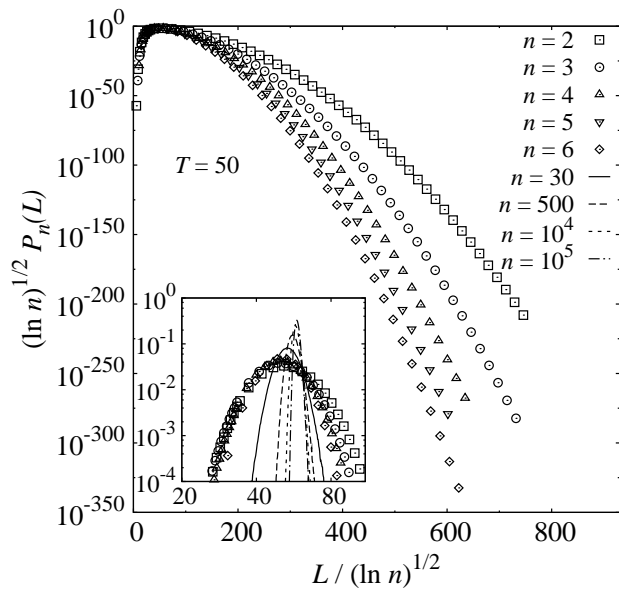


FIG. 12: Rescaled pdfs for  $T = 50$ , the perimeter  $L$  of the convex hull and various number of walks  $n$  in semi-logarithmic scale. Inset: Region close to peaks in double-logarithmic scale. Note that for walk numbers  $n \geq 30$  only values from simple sampling exist and therefore only the region around the peak is depicted.

$T = 10^4$  plotted in the same way as the rate function, also in double-logarithmic scale. Apparently, the shape of the rate functions shown in the main plot become more and more similar to the function shown in the inset. The actual values are quite different, because the values of  $n$  and  $T$  are very different for the two cases. This is due to strong corrections to the leading scaling behavior, as visible in Fig. 7, where also the minimum moves left and down when increasing the walk length  $T$ . Nevertheless, the result supports qualitatively the validity of Eq. (20). For a more quantitative statement the numbers  $n$  and  $T$  which can be studied using the large-deviation approach are too small due to the huge numerical effort which would be needed in this case.

Anyway, our results indicate that for  $n \rightarrow \infty$  and  $T \rightarrow \infty$  the distribution of the perimeter of many walkers can indeed be described by a suitably rescaled convolution of two Gumbel extreme-value distributions.

For the case of the area (not shown) we observe a similar behavior as for the perimeter. Again we found that close to the minimum a change of the shape starts to appear. Nevertheless, here we have no functional form available, so we do not discuss this further.

## V. CONCLUSION AND OUTLOOK

We have performed simulations of multiple two-dimensional discrete-time random walks with Gaussian displacements. Convex hulls of the random walks have been calculated and the area  $A$  and perimeter  $L$  have

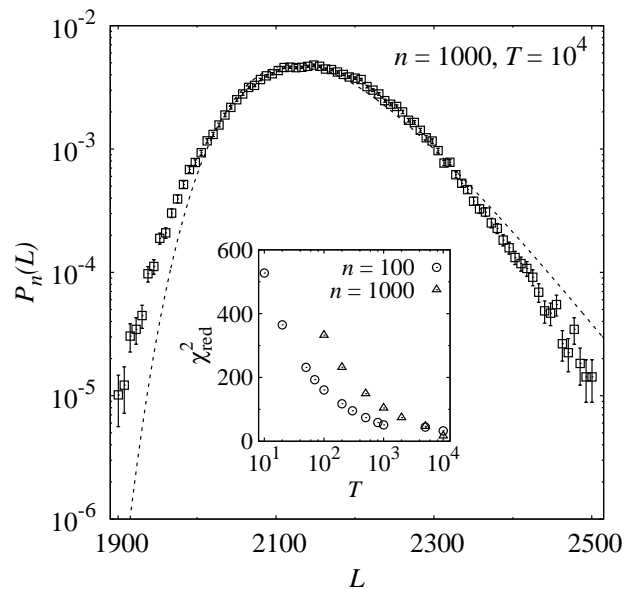


FIG. 13: Distribution of the perimeter for a larger number  $n = 1000$  of walkers and a very long walk length  $T = 10^4$ . The dashed line shows the result of a fit to the distribution given in Eq. (20), which resulted in  $a = 3.83 \cdot 10^{-3}$ ,  $B = 754.96$ , and  $C = 255.81$ , yielding  $\chi^2_{\text{red}} = 16.84$ . The inset shows the reduced chi-square for two different numbers of walkers as a function of the walk length.

been obtained. We have applied a large-deviation scheme, via biasing Markov-chain Monte Carlo evolutions in the configuration space of walks. The bias was introduced with respect to large or small areas or perimeters, respectively. In this way we have been able to obtain these distributions, for moderate number of walks, over large ranges of the support. Thus, we could measure probability densities spanning as many as 1000 decades in probability. The resulting probability densities show the same scaling behavior as the mean with respect to the length  $T$  of the walks, i.e.,  $\tilde{P}_n(A/T)$  and  $\tilde{P}_n(L/\sqrt{T})$  appear to be universal densities.

For small numbers  $n$  of walkers, the shape of these universal densities follows Gaussian distributions for  $L$  and  $\sqrt{A}$ , respectively, as for the  $n = 1$  case. Also, for the deviations of the distribution in the direction of very small diameters and areas, the previously found ( $n = 1$ ) essential singularity is obtained for low- $n$  multiple random walks.

We also obtained the rate functions for area and perimeter, rescaled with the scaling behavior of the maxima, i.e.,  $T^2$  and  $T$ , respectively. For both quantities, the finite-length rate functions approach limiting functions for  $T \rightarrow \infty$ , showing that the densities follow the large-deviation principle [51, 52]. This makes it likely that using analytical approaches from large-deviation theory, some results for the distributions of the convex hulls may be obtained. Anyway, the rate functions seem to be well described by a power law in the case  $n \rightarrow \infty$ , as found previously in the  $n = 1$  case.

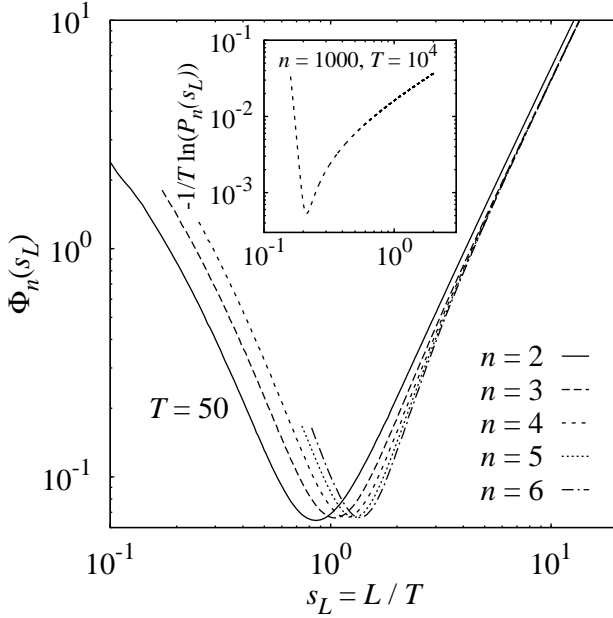


FIG. 14: Rate function  $\Phi_n(s_L)$  as a function of the scaled perimeter  $s_L = L/T$  for different number of walks  $n \in [2, 6]$  and walk length  $T = 50$  in double-logarithmic scale. Inset: the function from Eq. (20) with the parameters as obtained from the fit to the  $n = 1000$ ,  $T = 10^4$  data, rescaled as to obtain a rate function, also shown in double-logarithmic scale.

Finally, we have verified, that the scaling behavior of the averages with respect to the number  $n$  of walks is

predicted as in the literature [14]. The convergence is slow, such that on the level of the full distribution the convergence to a limiting function is not fully visible. Nevertheless, using simple-sampling simulations of number of walks up to  $n = 10^5$ , a convergence in the high-probability, i.e., peak region is visible, making a full convergence likely. Furthermore, these results are compatible with a convergence of the distribution of the perimeter to a convolution of two Gumbel extreme-value distributions.

For future research it would be interesting to investigate multiple interacting walkers [31, 39], or multiple walkers performing self-avoiding walks or loop-erased random walks [53, 54]. Furthermore, in higher dimensions a change of the scaling of the obtained distributions and thus also of the shape of the distributions can be anticipated, making such studies useful. Finally, it would be very interesting to apply the methods used here to biological models to investigate the formation of animal territories [9], which had originally motivated this work.

### Acknowledgments

This project was supported by the German Science Foundation (DPG) under grant HA 3169/8-1. The simulations were performed at the HERO cluster of the University of Oldenburg funded by the DFG (INST 184/108-1 FUGG) and the ministry of Science and Culture (MWK) of the Lower Saxony State.

- 
- [1] G. Polya, *Math. Ann.* **84**, 149 (1921).
  - [2] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1992).
  - [3] H. C. Berg, *Random Walks in Biology* (Princeton University Press, Princeton, 1993).
  - [4] B. D. Hughes, *Random walks and random environments* (Oxford University Press, Oxford, 1996).
  - [5] P. Bovet and S. Benhamou, *J. Theor. Biol.* **131**, 419 (1988).
  - [6] F. Bartumeus, M.G.E. Da Luz, G.M. Viswanathan and J. Catalan, *Ecology* **86**, 3078 (2005).
  - [7] C. O. Mohr, *American Midland Naturalist* **37**, 223 (1947).
  - [8] B. J. Worton, *Biometrics* **51**, 1206 (1995).
  - [9] L. Giuggioli, J. R. Potts, and S. Harris, *PLoS Comput. Biol.* **7**, e1002008 (2011).
  - [10] R. E. Kenward, *Wildlife Radio Tagging* (Academic Press, London, 1987).
  - [11] E. Landguth and F. Huettmann, *Spatial Complexity, Informatics, and Wildlife Conservation* (Springer, Heidelberg, 2010).
  - [12] B. J. Worton, *Ecol. Model.* **38**, 277 (1987).
  - [13] S. A. Boyle, W. C. Lourenco, L. R. da Silva, and A. T. Smith, *Folia Primatol.* **80**, 33 (2009).
  - [14] S. N. Majumdar, A. Comtet, and J. Randon-Furling, *J. Stat. Phys.* **138**, 955 (2010).
  - [15] L. Takács, *Amer. Math. Month.* **87**, 142 (1980).
  - [16] G. Letac, *J. Theor. Prob.* **6**, 385 (1993).
  - [17] M. El Bachir, *L'enveloppe convexe du mouvement brownien* (Ph.D thesis, Université Paul Sabatier, Toulouse, France, 1983).
  - [18] J. Randon-Furling, S. N. Majumdar, and A. Comtet, *Phys. Rev. Lett.* **103**, 140602 (2009).
  - [19] A. Reymbaut, S. N. Majumdar, and A. Rosso, *J. Phys. A: Math. Theor.* **44**, 415001 (2011).
  - [20] E. Dumonteil, S. N. Majumdar, A. Rosso, and A. Zoia, *Proc. Nat. Acad. Sci. USA*, **110**, 4239 (2013).
  - [21] M. Luković, T. Geisel, and S. Eule, *New. J. Phys.* **15**, 063034 (2013).
  - [22] M. Chupeau, O. Benichou, and S. N. Majumdar, *Phys. Rev. E* **91**, 050104(R) (2015).
  - [23] R. Eldan, *Electron. J. Probab.* **19**, no. 45, 1–34 (2014).
  - [24] Z. Kabluchko, and D. Zaporozhets, *arXiv:1404.6113*
  - [25] T. L. Snyder, and J. M. Steele, *Proc. Amer. Math. Soc.* **117**, 1165 (1993).
  - [26] A. Goldman, *Probab. Theor. Relat. Fields* **105**, 57 (1996).
  - [27] G. Claussen, A. K. Hartmann, and S. N. Majumdar, *Phys. Rev. E* **91**, 052104 (2015).
  - [28] H. Larralde, P. Trunfio, S. Havlin, H. E. Stanley, and

- G. H. Weiss, *Nature* **355**, 423 (1992).
- [29] H. Larralde, P. Trunfio, S. Havlin, H. E. Stanley, and G. H. Weiss, *Phys. Rev. A* **45**, 7128 (1992).
  - [30] S. B. Yuste, and L. Acedo, *Phys. Rev. E* **60**, R3459(R) (1999).
  - [31] L. Acedo and S. B. Yuste, *Recent Res. Devel. Stat. Phys.*, **2**, 83 (2002).
  - [32] P. J. Forrester, *J. Stat. Phys.* **56**, 767 (1989).
  - [33] V. Kukla, J. Kornatowski, D. Demuth, I. Girnus, H. Pfeifer, L. V. C. Rees, S. Schunk, K. K. Unger, and J. Kärger, *Science* **272**, 702 (1996).
  - [34] C. Aslangul, *J. Phys. A: Math. Gen.* **32**, 3993 (1999).
  - [35] G. Schehr, S. N. Majumdar, A. Comtet and J. Randon-Furling, *Phys. Rev. Lett.* **101**, 150601 (2008)
  - [36] P. J. Forrester, S. N. Majumdar and G. Schehr, *Nucl. Phys. B.* **844**, 500 (2011)
  - [37] J. Rambeau and G. Schehr, *Phys. Rev. E* **83**, 061146 (2011).
  - [38] G. Schehr, S. N. Majumdar, A. Comtet, P. J. Forrester, *J. Stat. Phys.* **150**, 491 (2013)
  - [39] A. Kundu, and S. N. Majumdar, and G. Schehr, *J. Stat. Phys.* **157**, 124 (2014).
  - [40] J. Dräger, and J. Klafter, *Phys. Rev. E* **60**, 6503 (1999).
  - [41] P. L. Krapivsky, S. N. Majumdar and A. Rosso, *Journal of Phys. A: Math. Theor.* **43**, 315001 (2010)
  - [42] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume I* (John Wiley & Sons, New York, 1950).
  - [43] F. P. Preparata, and M. I. Shamos, *Computational Geometry – An Introduction* (Springer, Berlin, 1985).
  - [44] R. A. Jarvis, *Inform. Process. Lett.* **2**, 18 (1973).
  - [45] S. G. Akl, and G. T. Toussaint, *Inform. Process. Lett.* **7**, 219 (1978).
  - [46] A. K. Hartmann, *Phys. Rev. E* **65**, 056102 (2002).
  - [47] A. K. Hartmann, *Eur. Phys. J. B* **84**, 627 (2011).
  - [48] A. K. Hartmann, *New Optimization Algorithms in Physics*, Editors: A. K. Hartmann, and H. Rieger (Wiley-VCH, Berlin, 2004; pp. 253).
  - [49] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
  - [50] A. Kundu, S.N. Majumdar and G. Schehr, *Phys. Rev. Lett.* **110**, 220602 (2013)
  - [51] F. denHollander, *Large Deviations* (American Mathematical Society, Providence, 2000).
  - [52] A. Dembo, and O. Zeitouni, *Large Deviation Techniques and Applications* (Springer, Berlin, 2010).
  - [53] G. F. Lawler, *J. Phys. A: Math. Gen.* **20**, 4565 (1987).
  - [54] S. N. Majumdar, *Phys. Rev. Lett.* **68**, 2329 (1992).